Contributions to spatial data analysis and Stein’s method

Marie Ernst

December 17th, 2019
Part I: Spatial dependence

1. Outliers in spatial multivariate data
2. Robustness of tests for spatial autocorrelation

Part II: Stein's method

3. Stein differentiation
4. First order covariance identities and inequalities
5. Infinite covariance expansions
6. Stein factors and distances between distributions
7. General conclusions and perspectives
Part 1: spatial dependence
Part 1: spatial dependence

Two research questions
Part 1: spatial dependence

Question 1
Spatial data:
- geographical positions
- non spatial attributes

Example
Waste per capita (kg) in the Walloon region in Belgium
Multivariate spatial data

Example 2D

Unemployment rate

Old buildings ($\geq 30$ years)
Multivariate spatial data

Example 2D

Spatial locations

Attribute representation
Multivariate spatial outliers

Two types of outliers (Haslett et al. (1991)):

- global outlier: extreme behaviour wrt all observations

Example 2D

Froidchapelle: global outlier
Multivariate spatial outliers

Two types of outliers (Haslett et al. (1991)):

- global outlier: extreme behaviour wrt all observations
- local outlier: extreme behaviour wrt its neighbours

Example 2D

Froidchapelle: global and local outlier
Multivariate spatial outliers

Two types of outliers (Haslett et al. (1991)):

- **global outlier**: extreme behaviour wrt all observations
- **local outlier**: extreme behaviour wrt its neighbours

Example 2D

Martelange: local outlier
Objectives in dimension $p$
Objectives in dimension $p$

Global outliers detection

- Geographical components not used
- Usual outlier detection techniques can be used
  $\Rightarrow$ not considered here
Objectives in dimension $p$

Global outliers detection
- Geographical components not used
- Usual outlier detection techniques can be used
  $\Rightarrow$ not considered here

Local outliers detection
- Review of some existing techniques
- Suggestion of an adaptation
- Comparison with examples and simulations
Considered Techniques

1. Chen et al. (2008)

2. Harris et al. (2014)

3. Filzmoser et al. (2014)
Considered Techniques

2. Harris *et al.* (2014)
3. Filzmoser *et al.* (2014)
Considered Techniques

1. Chen et al. (2008): Componentwise median and robust Mahalanobis distance
2. Harris et al. (2014): Geographically Weighted PCA with robust estimator
3. Filzmoser et al. (2014)
Considered Techniques

1. Chen et al. (2008): Componentwise median and robust Mahalanobis distance
2. Harris et al. (2014): Geographically Weighted PCA with robust estimator
3. Filzmoser et al. (2014): Robust “Mahalanobis-type” detection
Approach
Robust “Mahalanobis-type” detection

1 Preliminary global step:
Robust estimation of the general structure: \((\hat{\mu}, \hat{\Sigma})\)

Example 2D

Filzmoser, Ruiz-Gazen and Thomas-Agnan (2014)
Approach
Robust “Mahalanobis-type” detection

1. Preliminary global step:
   Robust estimation of the general structure: $\hat{(\mu, \Sigma)}$

2. Local step:

Example 2D
Approach

Robust “Mahalanobis-type” detection

1. Preliminary global step:
   Robust estimation of the general structure: \((\hat{\mu}, \hat{\Sigma})\)

2. Local step:
   • Centring the general structure on the observation

Example 2D

Unemployment rate

Proportion of old buildings
Filzmoser, Ruiz-Gazen and Thomas-Agnan (2014)

Approach
Robust “Mahalanobis-type” detection

1. Preliminary global step:
   Robust estimation of the general structure: \((\hat{\mu}, \hat{\Sigma})\)

2. Local step:
   - Centring the general structure on the observation
   - Determination of the ellipsoid containing the next neighbour

Example 2D

Unemployment rate
Proportion of old buildings

5 10 15 20 25
65 70 75 80 85 90 95

5 10 15 20 25
65 70 75 80 85 90 95

Unemployment rate
Proportion of old buildings
Approach

Robust “Mahalanobis-type” detection

1. Preliminary global step:
   Robust estimation of the general structure: $(\hat{\mu}, \hat{\Sigma})$

2. Local step:
   - Centring the general structure on the observation
   - Determination of the ellipsoid containing the next neighbour
   - If its tolerance level is larger than a theoretical quantile
     $\Rightarrow$ local outlier
Regularized spatial detection technique
(E. and Haesbroeck, 2017)

Approach: adaptation of Filzmoser et al. (2014)
Work with local structure and only on the most homogeneous neighbourhoods
Regularized spatial detection technique
(E. and Haesbroeck, 2017)

Approach: adaptation of Filzmoser et al. (2014)
Work with local structure and only on the most homogeneous neighbourhoods

1 Local step:
   • Estimation of the local structure: \((\hat{\mu}_i, \hat{\Sigma}_i)\) with robust and regularized estimators
Regularized spatial detection technique
(E. and Haesbroeck, 2017)

Approach: adaptation of Filzmoser et al. (2014)
Work with local structure and only on the most homogeneous neighbourhoods

1 Local step:
   • Estimation of the local structure: \((\hat{\mu}_i, \hat{\Sigma}_i)\) with robust and regularized estimators
   • Homogeneity measure: \(\text{det}(\hat{\Sigma}_i)\)
Regularized spatial detection technique
(E. and Haesbroeck, 2017)

Approach: adaptation of Filzmoser et al. (2014)

Work with local structure and only on the most homogeneous neighbourhoods

1 Local step:
   - Estimation of the local structure: \((\hat{\mu}_i, \hat{\Sigma}_i)\) with robust and regularized estimators
   - Homogeneity measure: \(\text{det}(\hat{\Sigma}_i)\)

2 Global step: Selection of 10\%, 20\%, \ldots of smallest values

3 Local step: work only on selected neighbourhoods
   - Centring the local structure on the observation
   - Determination of the ellipsoid containing the next neighbour
   - If its tolerance level is larger than an empirical quantile \(\Rightarrow\) local outlier
Illustration

Outliers for Filzmoser et al.

Outliers for regularization
Test on 10% of neighbourhoods
Outliers for Filzmoser et al.

Outliers for regularization
Test on 10% of neighbourhoods
Illustration

Outliers for Filzmoser et al.

Outliers for regularization
Test on 10% of neighbourhoods
Illustration

Outliers for Filzmoser et al. Outliers for regularization

Test on 20% of neighbourhoods
Illustration

Outliers for Filzmoser et al.

Outliers for regularization
Test on 30\% of neighbourhoods
Wallonia: 14 socio-economic variables for the 262 municipalities
Wallonia: 14 socio-economic variables for the 262 municipalities

Simulations for the comparison
Simulations

Generate spatial data of $p$ variables for $n$ locations (grid or Walloon municipalities)

Simulation set-up

- Matérn model to generate spatial data varying the overall smoothness
- Contamination by swapping observations with high/small PCA scores\(^1\)

\(^1\)Harris et al. (2014)
Results

• Harris et al. (2014) wrongly flags too many good observations as local outliers.

• Chen et al. (2008) handles well the regular domain with the less smooth design.

• Filzmoser et al (2014) outperforms the two previous techniques in most configurations (smoother variable and/or irregular domain).

• The adaptation has similar results as the initial technique; these results being dependent on the homogeneity constraint that we set.
Spatial autocorrelation

Question 2: When do we have to consider spatial techniques?
Question 2: When do we have to consider spatial techniques?
When the i.i.d. assumption is no longer satisfied?
Are the values randomly assigned to locations?
Spatial autocorrelation

Question 2: When do we have to consider spatial techniques?
When the i.i.d. assumption is no longer satisfied?
Are the values randomly assigned to locations?
⇒ check spatial autocorrelation
Spatial autocorrelation

Question 2: When do we have to consider spatial techniques?
When the i.i.d. assumption is no longer satisfied?
Are the values randomly assigned to locations?
⇝ check spatial autocorrelation

Positive spatial autocorrelation
Negative spatial autocorrelation
No spatial autocorrelation
Weighting matrix $W$

Locations $s_i$ and $s_j$ are neighbours if and only if $w_{ij} > 0$. Otherwise, $w_{ij} = 0$. 

Neighbours
Weighting matrix $W$

Locations $s_i$ and $s_j$ are neighbours if and only if $w_{ij} > 0$. Otherwise, $w_{ij} = 0$.

Different choices:

- Binary weights,
- Row-standardized,
- Globally standardized, . . .
Neighbours

Weighting matrix $W$

Locations $s_i$ and $s_j$ are neighbours if and only if $w_{ij} > 0$. Otherwise, $w_{ij} = 0$.

Different choices:

- Binary weights,
- Row-standardized,
- Globally standardized, . . .

Convention: zero diagonal and $S_0 = \sum_i \sum_j w_{ij}$.
Measures of spatial autocorrelation

Sample data points \( z = \{z_1, \ldots, z_n\} \) observed at spatial locations \( \{s_1, \ldots, s_n\} \)
Measures of spatial autocorrelation

Sample data points \( z = \{z_1, \ldots, z_n\} \) observed at spatial locations \( \{s_1, \ldots, s_n\} \)

Moran’s Index (1950)

\[
l(z) = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(z_i - \bar{z})(z_j - \bar{z})}{\sum_{i=1}^{n} (z_i - \bar{z})^2}
\]
Measures of spatial autocorrelation

Sample data points $\mathbf{z} = \{z_1, \ldots, z_n\}$ observed at spatial locations $\{s_1, \ldots, s_n\}$

Moran’s Index (1950)

$$I(\mathbf{z}) = \frac{n}{S_0} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (z_i - \bar{z})(z_j - \bar{z}) \frac{\sum_{i=1}^{n} (z_i - \bar{z})^2}{\sum_{i=1}^{n} (z_i - \bar{z})^2}$$
Measures of spatial autocorrelation

Sample data points \( z = \{ z_1, \ldots, z_n \} \) observed at spatial locations \( \{ s_1, \ldots, s_n \} \)

Moran’s Index (1950)

\[
I(z) = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\sum_{i=1}^{n} (z_i - \bar{z})^2}
\]

Geary’s ratio (1954)

\[
c(z) = \frac{n - 1}{2S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (z_i - z_j)^2}{\sum_{i=1}^{n} (z_i - \bar{z})^2}
\]

Getis and Ord’s statistics (1992)

\[
G(z) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} z_i z_j}{\sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} z_i z_j}
\]
Tests based on asymptotic normality

Without spatial autocorrelation, \( I, c \) and \( G \) are \textbf{asymptotically Gaussian} under normality (N) and/or randomisation (R) assumption.
Tests based on asymptotic normality

Without spatial autocorrelation, $I$, $c$ and $G$ are asymptotically **Gaussian** under normality (N) and/or randomisation (R) assumption.

<table>
<thead>
<tr>
<th>Moran’s $I$</th>
<th>Geary’s $c$</th>
<th>Getis and Ord’s $G$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test under R</td>
<td>Test under R</td>
<td>Test under R</td>
</tr>
<tr>
<td>Test under N</td>
<td>Test under N</td>
<td>Test under R</td>
</tr>
<tr>
<td>Permutation test</td>
<td>Permutation test</td>
<td>Permutation test</td>
</tr>
<tr>
<td>Dray’s test</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Robustness: example

Crude divorce rate in Belgium

- Crude divorce rate (2017): 4 per 1,000 population
- Crude divorce rate (2017): 21 per 1,000 population

- Moran's I = 0.01
- E[I] = −0.0017
- Range values: −0.67 ≤ I ≤ 1.16
- p-value = 0.25 (0.32) under R (under N)
Robustness: example

Crude divorce rate in Belgium

- Moran’s $I = 0.01$
- $E_{H_0}[I] = -0.0017$
- Range values: $-0.67 \leq I \leq 1.16$
- $p$-value $= 0.25$ (0.32) under R (under N)
Robustness of the tests

Robustness (Huber, 1981)
Insensitivity to small deviations from assumptions and more precisely, outlier resistance

Classic robustness tools
- Breakdown point (Hampel 1971)
- Influence function (Hampel et al. 1986)

Characteristic
Based on functionals \((n \to \infty)\)
Robustness of the tests

Robustness (Huber, 1981)
Insensitivity to small deviations from assumptions and more precisely, outlier resistance

Classic robustness tools
- Breakdown point (Hampel 1971)
- Influence function (Hampel et al. 1986)

Characteristic
Based on functionals \((n \to \infty) \rightsquigarrow\) favour empirical tools (finite \(n\))
Empirical influence function of the $p$-value

(Lambert 1981)

Definition

$$EIF(\xi, i) = \frac{p\text{-value}(z + \xi e_i) - p\text{-value}(z)}{1/n}$$
Empirical influence function of the $p$-value

(Lambert 1981)

Definition

$$EIF(\xi, i) = \frac{\text{p-value}(z + \xi e_i) - \text{p-value}(z)}{1/n}$$

Proposition (Chapter 2, Prop. 2.4.1)

The EIF of the $p$-value of unilateral tests based on asymptotic normality for Moran’s index is explicitly given by

$$EIF(\xi, i; I) = n \left[ \Phi \left( - \frac{I(z + \xi e_i) - \mathbb{E}[I]}{\sigma[I(z + \xi e_i)]} \right) - \Phi \left( - \frac{I(z) - \mathbb{E}[I]}{\sigma[I(z)]} \right) \right]$$

$$\rightarrow n \left[ \Phi \left( \frac{2(nw_i \cdot - S_0)}{(n - 1)S_0 \sigma[I]} \right) - \Phi \left( - \frac{I(z) - \mathbb{E}[I]}{\sigma[I]} \right) \right]$$

as $\xi$ tends to infinity.
Robustness: example

Crude divorce rate: hair-plot (Genton and Ruiz-Gazen, 2010)
Resistance of a test  
(Ylvisaker 1977)

Definition

Resistance to acceptance (resp. rejection): smallest proportion of the data that must be corrupted to guarantee the acceptance (rejection) of $H_0$. 

Proposition (Chapter 2, Prop. 2.4.2 - 2.4.3)

- Resistance to acceptance is $1/n$ for both asymptotic tests
- The resistance to rejection is $m/n$ where $m$ is the size of the smallest subset $A \subseteq \{1, \ldots, n\}$ which satisfies

$$n^2 w_A - 2mn (w_A + w_B) + m^2 S_0 S_0 m (n - m) > -n^{-1} + \sigma N [I] z_{1 - \alpha} \text{ under } N$$
Resistance of a test
(Ylvisaker 1977)

Definition

**Resistance to acceptance (resp. rejection):** smallest proportion of the data that must be corrupted to guarantee the acceptance (rejection) of $H_0$.

Proposition (Chapter 2, Prop. 2.4.2 - 2.4.3)

- Resistance to acceptance is $1/n$ for both asymptotic tests
- The resistance to rejection is $m/n$ where $m$ is the size of the smallest subset $A \subseteq \{1, \ldots, n\}$ which satisfies

\[
\begin{align*}
\frac{n^2 w_A - 2nm(w_A + w_B) + m^2 S_0}{S_0 m(n-m)} &> \frac{-1}{n-1} + \sigma_N[l] z_{1-\alpha} \quad \text{under } N \\
\frac{n^2 w_A - 2mn(w_A + w_B) + m^2 S_0}{S_0 m(n-m)} &> \frac{-1}{n-1} + \sqrt{a_1 - a_2 \frac{n^2 - 3nm + 3m}{(n-m)m}} z_{1-\alpha} \quad \text{under } R
\end{align*}
\]
Robustness: example

Example: Belgium

- *Resistance to acceptance*: for divorces, the contaminated rate of Brussels modified the result into an acceptance of $H_0$. If the “true” value is associated with Brussels\(^2\), Moran’s $I$ is 0.14 ($p$-value < 0.0001) instead of 0.01 ($p$-value $\geq$ 0.25).

\(^2\)Brussels: 24 local divorces vs 3698 divorces
Example: Belgium

- **Resistance to acceptance**: for divorces, the contaminated rate of Brussels modified the result into an acceptance of $H_0$. If the “true” value is associated with Brussels$^2$, Moran’s $I$ is 0.14 ($p$-value $< 0.0001$) instead of 0.01 ($p$-value $\geq 0.25$).

- **Resistance to rejection**: two corrupted neighbours are enough to always reject $H_0$.

$^2$Brussels: 24 local divorces vs 3698 divorces
Robustness: example

Example: Belgium

- **Resistance to acceptance**: for divorces, the contaminated rate of Brussels modified the result into an acceptance of $H_0$. If the “true” value is associated with Brussels\(^2\), Moran’s $I$ is $0.14$ ($p$-value $< 0.0001$) instead of $0.01$ ($p$-value $\geq 0.25$).

- **Resistance to rejection**: two corrupted neighbours are enough to always reject $H_0$.

**Other tests**
The lack of robustness of the other tests is similarly proved.

\(^2\)Brussels: 24 local divorces vs 3698 divorces
Robust alternative 1

Rank Moran index $I_r$

Idea: replace observations by their rank

Proposition (Chapter 2, Prop 2.5.2)

$$EIF(\xi, i; I_r) = n \left[ \Phi(I_r(z) - E[I_r]) - \Phi(I_r(z) - E[I_r]) + \sigma[I_r] \right]$$

where $\epsilon = I_r(z + \xi e_i) - I_r(z)$ is explicit.

The impact of contamination on the $p$-value is limited.

Estimated resistances (Belgium)

- Resistance to rejection: at most 14/589
- Resistance to acceptance: 6/589
Robust alternative 1

Rank Moran index $I_r$

**Idea**: replace observations by their rank

**Proposition** (Chapter 2, Prop 2.5.2)

$$\text{EIF}(\xi, i; I_r) = n \left[ \Phi \left( \frac{I_r(z) - E[I_r]}{\sigma[I_r]} \right) - \Phi \left( \frac{I_r(z) - E[I_r]}{\sigma[I_r]} + \frac{e}{\sigma[I_r]} \right) \right]$$

where $e = I_r(z + \xi e_i) - I_r(z)$ is explicit.

The impact of contamination on the $p$-value is limited.
Robust alternative 1

Rank Moran index $I_r$

**Idea:** replace observations by their rank

**Proposition (Chapter 2, Prop 2.5.2)**

$$EIF(\xi, i; I_r) = n \left[ \Phi \left( \frac{I_r(z) - E[I_r]}{\sigma[I_r]} \right) - \Phi \left( \frac{I_r(z) - E[I_r]}{\sigma[I_r]} + \frac{e}{\sigma[I_r]} \right) \right]$$

where $e = I_r(z + \xi e_i) - I_r(z)$ is explicit.

The impact of contamination on the $p$-value is limited.

**Estimated resistances (Belgium)**

- Resistance to rejection: at most 14/589
- Resistance to acceptance: 6/589
Moran scatterplot

Moran can be interpreted as the slope in a OLS regression of spatially lagged observations over $z$. 
Robust alternative 2

Moran scatterplot

Moran can be interpreted as the slope in a OLS regression of spatially lagged observations over $z$.

$\Rightarrow$ **Idea:** use robust regression
Robust and efficient regression estimation

Different methods

• S-estimator (Rousseeuw and Yohai, 1984)
• Least Trimmed Squares (LTS - Rousseeuw, 1985)
• MM-estimator (Yohai, 1987)
• Robust and Efficient Weighted Least Squares Estimator (REWLSE - Gervini and Yohai, 2002)
Robust and efficient regression estimation

Different methods

- S-estimator (Rousseeuw and Yohai, 1984)
- Least Trimmed Squares (LTS - Rousseeuw, 1985)
- MM-estimator (Yohai, 1987)
- Robust and Efficient Weighted Least Squares Estimator (REWLSE - Gervini and Yohai, 2002)

Robustness of adapted Moran index

Due to robust properties of regression estimators,

- the impact on the $p$-value is zero almost everywhere;
- the resistance is $m/n$ where $m$ is the minimal number of observations for which the joint neighbourhoods contains at most 50% of the points in Moran scatterplot.
Simulation study

Efficiency of robust tests
Comparison of level and power of robust and classic tests.

Spatial autoregressive model
\[ Z = \rho W Z + \varepsilon \]
where \( \rho \) is the spatial correlation coefficient.

Results
• Comparable power and level for all tests
• Power increases with \( n \)
• Power increases with \( \rho \)
Simulation study

Efficiency of robust tests
Comparison of level and power of robust and classic tests.

Spatial autoregressive model

\[ Z = \rho WZ + \varepsilon \]

where \( \rho \) is the spatial correlation coefficient.
Efficiency of robust tests
Comparison of level and power of robust and classic tests.

Spatial autoregressive model

\[ Z = \rho W Z + \varepsilon \]

where \( \rho \) is the spatial correlation coefficient.

Results

- Comparable power and level for all tests
- Power increases with \( n \)
- Power increases with \( \rho \)
End of Part 1
Part 2: Stein’s method
Stein’s identity

If $X \sim \mathcal{N}(0, 1)$, then

$$
\mathbb{E}[Xg(X)] = \mathbb{E}[g'(X)] \quad \forall g \text{ s.t. } \mathbb{E}[|g'(X)|] < \infty
$$
Stein’s identity
If $X \sim \mathcal{N}(0, 1)$, then

$$\mathbb{E}[Xg(X)] = \mathbb{E}[g'(X)] \quad \forall g \text{ s.t. } E[|g'(X)|] < \infty$$

Stein operator
The associated Stein operator is defined by

$$\mathcal{T}g(x) = xg(x) - g'(x)$$

and $\mathbb{E}[\mathcal{T}g(X)] = 0$ for any appropriate $g$. 

Gaussian case
Formalism

Let $X \sim p$. The derivative-type operators are

$$\Delta^\ell f(x) = \begin{cases} f'(x) & \text{if } \ell = 0; \\ f(x + 1) - f(x) & \text{if } \ell = +1; \\ f(x) - f(x - 1) & \text{if } \ell = -1. \end{cases}$$
Let $X \sim p$. The derivative-type operators are

$$
\Delta^\ell f(x) = \begin{cases} 
    f'(x) & \text{if } \ell = 0; \\
    f(x + 1) - f(x) & \text{if } \ell = +1; \\
    f(x) - f(x - 1) & \text{if } \ell = -1.
\end{cases}
$$

**Canonical Stein operator**

For any $f : \mathbb{R} \to \mathbb{R}$, the *canonical Stein operator* is defined as

$$
\mathcal{T}^\ell_p : f(x) \mapsto \mathcal{T}^\ell_p f(x) := \frac{\Delta^\ell(f(x)p(x))}{p(x)}.
$$

For any $f$ from the *canonical Stein class*, $\mathbb{E}[\mathcal{T}_p f(X)] = 0$. 
Let $X \sim p$. The derivative-type operators are

$$\Delta^\ell f(x) = \begin{cases} 
  f'(x) & \text{if } \ell = 0; \\
  f(x+1) - f(x) & \text{if } \ell = +1; \\
  f(x) - f(x-1) & \text{if } \ell = -1.
\end{cases}$$

**Canonical Stein operator**

For any $f : \mathbb{R} \to \mathbb{R}$, the *canonical Stein operator* is defined as

$$\mathcal{T}_p^\ell : f(x) \mapsto \mathcal{T}_p^\ell f(x) := \frac{\Delta^\ell (f(x)p(x))}{p(x)}.$$ 

For any $f$ from the *canonical Stein class*, $\mathbb{E}[\mathcal{T}_p f(X)] = 0$.

**Particular case: the score function**

$$\rho_p^\ell(x) = \mathcal{T}_p^\ell 1(x) = \frac{\Delta^\ell p(x)}{p(x)}.$$
Example: Poisson density

\[ p(x) = e^{-\lambda} \frac{\lambda^x}{x!} \text{ for } x \in \mathbb{N} \text{ and } \ell = \pm 1 \]
Example: Poisson density

\[ p(x) = e^{-\lambda} \lambda^x / x! \text{ for } x \in \mathbb{N} \text{ and } \ell = \pm 1 \]

Canonical Stein operators

- \( \mathcal{T}_p^+ f(x) = \frac{\Delta^+(f(x)p(x))}{p(x)} = f(x+1)\frac{\lambda}{x+1} - f(x) \)
- \( \mathcal{T}_p^- f(x) = \frac{\Delta^-(f(x)p(x))}{p(x)} = f(x) - f(x-1)\frac{x}{\lambda} \)
Example: Poisson density

\[ p(x) = e^{-\lambda} \frac{\lambda^x}{x!} \text{ for } x \in \mathbb{N} \text{ and } \ell = \pm 1 \]

Canonical Stein operators

- \( \mathcal{T}_p^+ f(x) = \frac{\Delta^+(f(x)p(x))}{p(x)} = f(x+1) \frac{\lambda}{x+1} - f(x) \)
- \( \mathcal{T}_p^- f(x) = \frac{\Delta^-(f(x)p(x))}{p(x)} = f(x) - f(x-1) \frac{x}{\lambda} \)

The score functions

- \( \rho_p^+(x) = \mathcal{T}_p^+ 1(x) = \frac{\lambda}{x+1} - 1 \)
- \( \rho_p^- f(x) = \mathcal{T}_p^- 1(x) = 1 - \frac{x}{\lambda} \)
Pseudo inverse Stein operator

The canonical pseudo inverse Stein operator for the operator $\mathcal{T}_p^{\ell}$ is

$$\mathcal{L}_p^{\ell} : h(x) \mapsto \mathcal{L}_p^{\ell} h(x) := \frac{1}{p(x)} \int_{a}^{x-a_{\ell}} (h(u) - \mathbb{E}[h(X)]) p(u) \mu(du)$$

where $a_{\ell} = \mathbb{I}[\ell = 1]$.

Properties

- $\mathcal{L}_p^{\ell} \mathcal{T}_p^{\ell} f(x) = f(x)$
- $\mathcal{T}_p^{\ell} \mathcal{L}_p^{\ell} h(x) = h(x) - \mathbb{E}[h(X)]$
Pseudo inverse Stein operator

The canonical pseudo inverse Stein operator for the operator \( \mathcal{T}_p^\ell \) is

\[
\mathcal{L}_p^\ell : h(x) \mapsto \mathcal{L}_p^\ell h(x) := \frac{1}{p(x)} \int_a^{x-a_\ell} (h(u) - \mathbb{E}[h(X)]) p(u) \mu(du)
\]

where \( a_\ell = \mathbb{I}[\ell = 1] \).

Properties

- \( \mathcal{L}_p^\ell \mathcal{T}_p^\ell f(x) = f(x) \)
- \( \mathcal{T}_p^\ell \mathcal{L}_p^\ell h(x) = h(x) - \mathbb{E}[h(X)] \)

Particular case: the Stein kernel

\[
\tau_p^\ell(x) = -\mathcal{L}_p^\ell \text{Id}(x)
\]
Pseudo inverse Stein operator

The canonical pseudo inverse Stein operator for the operator $\mathcal{T}_p^\ell$ is

$$\mathcal{L}_p^\ell : h(x) \mapsto \mathcal{L}_p^\ell h(x) := \frac{1}{p(x)} \int_{a}^{x-a} (h(u) - \mathbb{E}[h(X)]) p(u) \mu(du)$$

where $a_\ell = \mathbb{I}[\ell = 1]$.

Properties

- $\mathcal{L}_p^\ell \mathcal{T}_p^\ell f(x) = f(x)$
- $\mathcal{T}_p^\ell \mathcal{L}_p^\ell h(x) = h(x) - \mathbb{E}[h(X)]$

Particular case: the Stein kernel

$$\tau_p^\ell(x) = -\mathcal{L}_p^\ell \text{Id}(x)$$

Poisson density

- $\tau_p^+(x) = x$ and $\tau_p^-(x) = \lambda$
Stein equation

Standardized Stein operator

\[ \mathcal{A}g(x) = T^\ell_p \left( f(\cdot)g(\cdot - \ell) \right)(x) = T^\ell_p f(x)g(x) + f(x)\Delta^{-\ell} g(x) \]
Stein equation

Standardized Stein operator

\[ \mathcal{A}g(x) = T_p^{\ell} (f(\cdot)g(\cdot - \ell))(x) = T_p^{\ell} f(x)g(x) + f(x)\Delta^{-\ell} g(x) \]

Stein covariance identity (IBP 1)

For all “appropriate” \( f \) and \( g \),

\[ \mathbb{E} \left[ \left( T_p^{\ell} f(X) \right) g(X) \right] = -\mathbb{E} \left[ f(X)\Delta^{-\ell} g(X) \right] \]
Stein equation

Standardized Stein operator

\[ A g(x) = T^\ell_p (f(\cdot) g(\cdot - \ell))(x) = T^\ell_p f(x) g(x) + f(x) \Delta^{-\ell} g(x) \]

Stein covariance identity (IBP 1)

For all “appropriate” \( f \) and \( g \),

\[ \mathbb{E} \left[ (T^\ell_p f(X)) g(X) \right] = - \mathbb{E} \left[ f(X) \Delta^{-\ell} g(X) \right] \]

Stein covariance identity (IBP 2)

For all “appropriate” \( f \) and \( g \),

\[ \text{Cov} [f(X), g(X)] = \mathbb{E} \left[ - \left( \mathcal{L}^\ell_p f(X) \right) \Delta^{-\ell} g(X) \right] \]
Representations of the inverse operator
E., Reinert and Swan (2019)

\[ \mathcal{L}_p^\ell f(x) = \mathbb{E} \left[ (f(X) - \mathbb{E}[f(X)]) \frac{\mathbb{I}[X \leq x - a_{\ell}]}{p(x)} \right] \] (Definition)
Representations of the inverse operator

E., Reinert and Swan (2019)

- \( \mathcal{L}_p^\ell f(x) = \mathbb{E} \left[ (f(X) - \mathbb{E}[f(X)]) \frac{\mathbb{I}[X \leq x - a_\ell]}{p(x)} \right] \) (Definition)
- \( -\mathcal{L}_p^\ell f(x) = \mathbb{E} \left[ (f(X_2) - f(X_1))\Phi(X_1, x, X_2) \right] \) (Rep. I)

where \( \Phi(u, x, v) = \mathbb{I}[u + a_\ell \leq x \leq v - b_\ell]/p(x) \)
Representations of the inverse operator
E., Reinert and Swan (2019)

\[ \mathcal{L}_p^\ell f(x) = \mathbb{E} \left[ (f(X) - \mathbb{E}[f(X)]) \frac{\mathbb{I}[X \leq x - a_\ell]}{p(x)} \right] \] (Definition)

\[ -\mathcal{L}_p^\ell f(x) = \mathbb{E} \left[ (f(X_2) - f(X_1)) \Phi(X_1, x, X_2) \right] \] (Rep. I)
where \( \Phi(u, x, v) = \mathbb{I}[u + a_\ell \leq x \leq v - b_\ell]/p(x) \)

\[ -\mathcal{L}_p^\ell f(x) = \mathbb{E} \left[ \frac{K_p^\ell(X, x)}{p(X)p(x)} \Delta^{-\ell} f(X) \right] \] (Rep. II)
where \( K_p^\ell(x, x') = \mathbb{P}[X \leq (x \wedge x') - a_\ell] \mathbb{P}[X \geq (x \vee x') + b_\ell] \)
Representations of the inverse operator
E., Reinert and Swan (2019)

\[ \mathcal{L}_p^\ell f(x) = \mathbb{E} \left[ \left( f(X) - \mathbb{E}[f(X)] \right) \frac{\mathbb{I}[X \leq x - a_\ell]}{p(x)} \right] \]  
(Definition)

\[ -\mathcal{L}_p^\ell f(x) = \mathbb{E} \left[ (f(X_2) - f(X_1))\Phi(X_1, x, X_2) \right] \]  
(Rep. I)

where \( \Phi(u, x, v) = \mathbb{I}[u + a_\ell \leq x \leq v - b_\ell]/p(x) \)

\[ -\mathcal{L}_p^\ell f(x) = \mathbb{E} \left[ \frac{K_p^\ell(X, x)}{p(X)p(x)} \Delta^{-\ell} f(X) \right] \]  
(Rep. II)

where \( K_p^\ell(x, x') = \mathbb{P}[X \leq (x \land x') - a_\ell] \mathbb{P}[X \geq (x \lor x') + b_\ell] \)

References for \( \ell = 0 \):

- Rep. I is non-explicitly given in Chatterjee and Shao (2011)
- Rep. II is available in Saumard (2019)
- Symmetric kernel \( K_p^0 \): first appearance attributed to Höffding (1940)
Examples

\[ x \mapsto \frac{K_p^l(x, x')}{p(x)} \]

\[ N(0, 1) \]

\[ N(\mu, \sigma^2) \]

\[ \text{Bin}(50, 0.2) \]

\[ \text{Bin}(50, \theta) \]
Applications

Application 1: First order variance bounds
Family of first order lower and upper covariance bounds for functionals of arbitrary univariate distributions.
⇝ E., Reinert and Swan (2019a), Bernoulli.

Application 2: Infinite covariance expansions
Covariance expansions for functionals of arbitrary univariate distributions.
⇝ E., Reinert and Swan (2019b), arXiv:1906.08376

Application 3: Stein factor and distances
Estimates of distances between univariate probability distributions.
Applications

Application 1: First order variance bounds
Family of first order lower and upper covariance bounds for functionals of arbitrary univariate distributions.
⇝ E., Reinert and Swan (2019a), Bernoulli.
Applications

Application 1: First order variance bounds
Family of first order lower and upper covariance bounds for functionals of arbitrary univariate distributions.
⇝ E., Reinert and Swan (2019a), Bernoulli.

Application 2: Infinite covariance expansions
Covariance expansions for functionals of arbitrary univariate distributions.
⇝ E., Reinert and Swan (2019b), arXiv:1906.08376
Application 1: First order variance bounds
Family of first order lower and upper covariance bounds for functionals of arbitrary univariate distributions.
⇝ E., Reinert and Swan (2019a), Bernoulli.

Application 2: Infinite covariance expansions
Covariance expansions for functionals of arbitrary univariate distributions.
⇝ E., Reinert and Swan (2019b), arXiv:1906.08376

Application 3: Stein factor and distances
Estimates of distances between univariate probability distributions.
Application 1: first order variance bounds

Theorem (Cramer-Rao and weighted Poincaré inequality)

\[
\mathbb{E} \left[ -\mathcal{L}_p^\ell h(X) (\Delta^{-\ell} g(X)) \right]^2 \leq \text{Var}[g(X)] \leq \mathbb{E} \left[ (\Delta^{-\ell} g(X))^2 \frac{-\mathcal{L}_p^\ell h(X)}{\Delta^{-\ell} h(X)} \right]
\]

for decreasing \( h \) with equality if and only if \( g \propto h \).
Application 1: first order variance bounds

Theorem (Cramer-Rao and weighted Poincaré inequality)

\[
\mathbb{E} \left[ -\mathcal{L}_p^\ell h(X)(\Delta^{-\ell} g(X)) \right]^2 \leq \text{Var}[g(X)] \leq \mathbb{E} \left[ (\Delta^{-\ell} g(X))^2 - \mathcal{L}_p^\ell h(X) \Delta^{-\ell} h(X) \right]
\]

for decreasing \( h \) with equality if and only if \( g \propto h \).

Examples

- (Chernoff, 1980) If \( X \sim \mathcal{N}(0, 1) \) then

\[
\mathbb{E}[g'(X)]^2 \leq \text{Var}[g(X)] \leq \mathbb{E}[(g'(X))^2]
\]

- (Brascamp-Lieb, 1976) For appropriate density and \( g \),

\[
\frac{\mathbb{E} \left[ (\Delta^{-\ell} g(X))^2 \right]}{\mathbb{E} \left[ (\rho_p^\ell(X))^2 \right]} \leq \text{Var}[g(X)] \leq \mathbb{E} \left[ (\Delta^{-\ell} g(X))^2 \right] - \Delta^{-\ell} \rho_p^\ell(X)
\]
Application 2: infinite expansion

\[ \text{Cov}[f(X)] = \sum_{k=1}^{n} (-1)^{k-1} E[\Delta - \ell f_{k-1}(X) \Delta - \ell f'_k(X)] + (-1)^n \text{for weight sequences } \Gamma_{\ell h}\] and remainder term \( R_{\ell h n}(h). \)

References:
- \( h_k(x) = x: \) Papathanasiou (1988, \( \ell k = 0 \))
- \( h_k(x) = x: \) Afendras et al. (2007, \( \ell k = -1 \))
- \( n = 1 \) and \( h_1 = -\log(p)' \) for log-concave \( p: \) Brascamp-Lieb inequality
Application 2: infinite expansion

Papathanasiou-type expansion

\[ \text{Cov} \left[ f(X) \right] = \sum_{k=1}^{n} (-1)^{k-1} \mathbb{E} \left[ \Delta^{-\ell_k} f_{k-1}(X) \Delta^{-\ell_k} f'_{k-1}(X) \frac{\Gamma_{k}^\ell h(X)}{\Delta^{-\ell_k} h_k(X)} \right] \]

\[ + \ (-1)^n R^\ell_n(h) \]

for weight sequences \( \Gamma_{k}^\ell h(x) \) and remainder term \( R^\ell_n(h) \).
Application 2: infinite expansion

Papathanasiou-type expansion

\[
\text{Cov} [f(X)] = \sum_{k=1}^{n} (-1)^{k-1} \mathbb{E} \left[ \Delta^{-\ell_k} f_{k-1}(X) \Delta^{-\ell_k} f'_{k-1}(X) \frac{\Gamma^\ell_k h(X)}{\Delta^{-\ell_k} h_k(X)} \right] \\
+ (-1)^n R_n^\ell(h)
\]

for weight sequences \( \Gamma^\ell_k h(x) \) and remainder term \( R_n^\ell(h) \).
Application 2: infinite expansion

Papathanasiou-type expansion

\[ \text{Cov} \left[ f(X) \right] = \sum_{k=1}^{n} (-1)^{k-1} \mathbb{E} \left[ \Delta^{-\ell_k} f_{k-1}(X) \Delta^{-\ell_k} f'_{k-1}(X) \frac{\Gamma_{\ell_k} h(X)}{\Delta^{-\ell_k} h_k(X)} \right] \]

\[ + (-1)^n R_{n}^{\ell}(h) \]

for weight sequences \( \Gamma_{\ell_k} h(x) \) and remainder term \( R_{n}^{\ell}(h) \).

References:

- \( h_k(x) = x \): Papathanasiou (1988, \( \ell_k = 0 \))
- \( h_k(x) = x \): Afendras et al. (2007, \( \ell_k = -1 \))
- \( n = 1 \) and \( h_1 = -(\log p)' \) for log-concave \( p \): Brascamp-Lieb inequality
Examples

Normal expansion

if $X \sim \mathcal{N}(0, 1)$ then $\Gamma_k^0(x) = \frac{1}{k!}$ for all $k$ and

$$\text{Cov}[f(X), g(X)] = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k!} \mathbb{E} \left[ f^{(k)}(X)g^{(k)}(X) \right].$$
Examples

Normal expansion

if $X \sim \mathcal{N}(0, 1)$ then $\Gamma^0_k(x) = \frac{1}{k!}$ for all $k$ and

$$\text{Cov}[f(X), g(X)] = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k!} \mathbb{E} \left[ f^{(k)}(X)g^{(k)}(X) \right].$$

Beta expansion

If $X \sim \text{Beta}(a, b)$ then $\Gamma^0_k(x) = \frac{(x(1-x))^k}{k!(a+b)[k]}$ for all $k$ and

$$\text{Cov}[f(X), g(X)] = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k!(a+b)[k]} \mathbb{E} \left[ f^{(k)}(X)g^{(k)}(X)X^k(1 - X)^k \right].$$
Poisson expansion

If $X \sim \text{Poi}(\lambda)$, the order 1 expansions are

$$\text{Var}[g(X)] = \mathbb{E} \left[ X(\Delta^- g(X))^2 \right] - R_1^+$$

$$= \lambda \mathbb{E} \left[ (\Delta^+ g(X))^2 \right] - R_1^-;$$
Examples (2)

**Poisson expansion**

If $X \sim \text{Poi}(\lambda)$, the order 1 expansions are

$$\text{Var}[g(X)] = \mathbb{E} [X(\Delta^- g(X))^2] - R_1^+$$

$$= \lambda \mathbb{E} [(\Delta^+ g(X))^2] - R_1^-;$$

and the order 2 expansions are

$$\text{Var}[g(X)] = \mathbb{E} [X(\Delta^- g(X))^2] - \frac{1}{2} \mathbb{E} [X(X - 1)(\Delta^{--} g(X))^2] - R_2^{++}$$

$$= \mathbb{E} [X(\Delta^- g(X))^2] - \frac{1}{2} \lambda \mathbb{E} [X(\Delta^{+-} g(X))^2] - R_2^{+-}$$

$$= \lambda \mathbb{E} [(\Delta^+ g(X))^2] - \frac{1}{2} \mathbb{E} [X(\Delta^{+-} g(X))^2] - R_2^{--}$$

$$= \lambda \mathbb{E} [(\Delta^+ g(X))^2] - \frac{1}{2} \lambda^2 \mathbb{E} [(\Delta^{++} g(X))^2] - R_2^{--}$$
Application 3: Stein factors and distances

Let $X_n$ and $X_\infty$ be random variables.
Application 3: Stein factors and distances

Let $X_n$ and $X_\infty$ be random variables.

Stein identity (IBP 1)

Two linear operators $\mathcal{T}_\infty^\ell$ and $\mathcal{L}_\infty^\ell$ are associated with $X_\infty$ such that

$$
\mathbb{E} \left[ (\mathcal{T}_\infty^\ell c(X_\infty)) g(X_\infty) \right] = -\mathbb{E} \left[ c(X_\infty) \Delta^{-\ell} g(X_\infty) \right]
$$

are valid for all sufficiently regular functions $c, g$. 

Stein discrepancy

Quantify the “distance” between the laws of $X_n$ and $X_\infty$: 

$$
S(X_n, X_\infty, G) := \sup_{g \in G} \left| \mathbb{E} \left[ (\mathcal{T}_\infty^\ell c(X_\infty)) g(X_\infty) + c(X_\infty) \Delta^{-\ell} g(X_\infty) \right] \right|
$$

Application 3: Stein factors and distances

Let $X_n$ and $X_\infty$ be random variables.

Stein identity (IBP 1)

Two linear operators $T^\ell_\infty$ and $L^\ell_\infty$ are associated with $X_\infty$ such that

$$
\mathbb{E} \left[ (T^\ell_\infty c(X_\infty)) g(X_\infty) \right] = -\mathbb{E} \left[ c(X_\infty) \Delta^{-\ell} g(X_\infty) \right]
$$

are valid for all sufficiently regular functions $c, g$.

In particular, if $X_n \not\overset{\mathcal{L}}{=} X_\infty$, there exist some $g$ functions such that

$$
\mathbb{E} \left[ (T^\ell_\infty c(X_n)) g(X_n) \right] \neq -\mathbb{E} \left[ c(X_n) \Delta^{-\ell} g(X_n) \right].
$$
Application 3: Stein factors and distances

Let $X_n$ and $X_\infty$ be random variables.

Stein identity (IBP 1)

Two linear operators $\mathcal{T}_\infty^\ell$ and $\mathcal{L}_\infty^\ell$ are associated with $X_\infty$ such that

$$E \left[ (\mathcal{T}_\infty^\ell c(X_\infty)) g(X_\infty) \right] = -E \left[ c(X_\infty) \Delta^{-\ell} g(X_\infty) \right]$$

are valid for all sufficiently regular functions $c, g$.

In particular, if $X_n \not\sim X_\infty$, there exist some $g$ functions such that

$$E \left[ (\mathcal{T}_\infty^\ell c(X_n)) g(X_n) \right] \neq -E \left[ c(X_n) \Delta^{-\ell} g(X_n) \right].$$

Stein discrepancy

Quantify the “distance” between the laws of $X_n$ and $X_\infty$:

$$S(X_n, X_\infty, \mathcal{G}) := \sup_{g \in \mathcal{G}} \left| E \left[ (\mathcal{T}_\infty^\ell c(X_n)) g(X_n) + c(X_n) \Delta^{-\ell} g(X_n) \right] \right|$$
Integral Probability Metric (IPM)

\[ D_{\mathcal{H}}(X_n, X_\infty) = \sup_{h \in \mathcal{H}} |\mathbb{E} h(X_n) - \mathbb{E} h(X_\infty)| \]
Integral Probability Metric (IPM)

\[ D_{\mathcal{H}}(X_n, X_\infty) = \sup_{h \in \mathcal{H}} |\mathbb{E}h(X_n) - \mathbb{E}h(X_\infty)| \]

Examples

- **Kolmogorov distance**: \( \sup_{z \in \mathbb{R}} |\mathbb{P}(X_n \leq z) - \mathbb{P}(X_\infty \leq z)| \)

  is associated with \( \mathcal{H}_{\text{Kol}} = \{ h(x) = \mathbb{I}[x \in (-\infty, z]) : z \in \mathbb{R} \} \).
Integral Probability Metric (IPM)

\[ \mathcal{D}_\mathcal{H}(X_n, X_\infty) = \sup_{h \in \mathcal{H}} |\mathbb{E}h(X_n) - \mathbb{E}h(X_\infty)| \]

Examples

- **Kolmogorov distance:** \( \sup_{z \in \mathbb{R}} |\mathbb{P}(X_n \leq z) - \mathbb{P}(X_\infty \leq z)| \)
  is associated with \( \mathcal{H}_{\text{Kol}} = \{ h(x) = \mathbb{I}[x \in (-\infty, z]) : z \in \mathbb{R} \} \).

- **Total variation distance:** \( \sup_{B \subset \mathbb{R}} |\mathbb{P}(X_n \in B) - \mathbb{P}(X_\infty \in B)| \)
  is associated with \( \mathcal{H}_{\text{TV}} = \{ h(x) = \mathbb{I}[x \in B] : B \in \mathcal{B}(\mathbb{R}) \} \).
IPM and Stein discrepancy

Integral Probability Metric (IPM)

\[ \mathcal{D}_\mathcal{H}(X_n, X_\infty) = \sup_{h \in \mathcal{H}} |\mathbb{E}h(X_n) - \mathbb{E}h(X_\infty)| \]

Examples

• **Kolmogorov distance:** \( \sup_{z \in \mathbb{R}} |\mathbb{P}(X_n \leq z) - \mathbb{P}(X_\infty \leq z)| \)
  is associated with \( \mathcal{H}_{\text{Kol}} = \{ h(x) = \mathbb{I}[x \in (-\infty, z]] : z \in \mathbb{R} \} \).

• **Total variation distance:** \( \sup_{B \subset \mathbb{R}} |\mathbb{P}(X_n \in B) - \mathbb{P}(X_\infty \in B)| \)
  is associated with \( \mathcal{H}_{\text{TV}} = \{ h(x) = \mathbb{I}[x \in B] : B \in \mathcal{B}(\mathbb{R}) \} \).

• **Wasserstein distance:** \( \int_{-\infty}^{\infty} |\mathbb{P}(X_n \leq z) - \mathbb{P}(X_\infty \leq z)| \, dz \)
  is associated with \( \mathcal{H}_{\text{Wass}} = \{ h : |h(x) - h(y)| \leq |x - y| \} \).
IPM and Stein discrepancy

Integral Probability Metric (IPM)

\[ D_{\mathcal{H}}(X_n, X_\infty) = \sup_{h \in \mathcal{H}} |E h(X_n) - E h(X_\infty)| \]

Theorem

\[ D_{\mathcal{H}}(X_n, X_\infty) = S(X_n, X_\infty, \mathcal{G}_{\mathcal{H}}) \]

where

\[ \mathcal{G}_{\mathcal{H}} = \left\{ g_{h}(x) = \frac{L_{p} h(x + \ell)}{c(x + \ell)} : h \in \mathcal{H} \right\}. \]
Examples

Lower half-line indicator: $h(x) = \mathbb{I}[x \leq \xi] \ (\ell = 0)$

- $g(x) = \frac{1}{c(x)} \frac{P(\bar{\xi} \land x) \bar{P}(\bar{\xi} \lor x)}{p(x)}$

- $g'(x) = \frac{\mathbb{I}[x \leq \xi] - P(\xi)}{c(x)} - \frac{T^0_p c(x)}{c^2(x)} \frac{P(\bar{\xi} \land x) \bar{P}(\bar{\xi} \lor x)}{p(x)}$.

$g(x)$ and $|g'(x)|$ for Gaussian target
Examples

Point mass: $h(x) = \mathbb{I}[x = \xi] \ (\ell = \pm 1)$

- $g(x) = \frac{p(\xi)}{c(x+\ell)p(x+\ell)} \left( \mathbb{I}[x \geq \xi + b_\ell] - P(x - b_\ell) \right)$
- If $c = \tau^\ell_p$,
  \[ \Delta^{-\ell} g_\xi^\ell(x) = \frac{\mathbb{I}[x=\xi]-p(\xi)}{\tau^-_p(x)} + \frac{p(\xi)(\mathbb{I}[x\geq\xi]-P(x))}{p(x)} \left( \frac{1}{\tau^-_p(x)} - \frac{1}{\tau^+_p(x)} \right) \]

$g(x)$ and $|\Delta^{-} g(x)|$ for Poisson target ($\ell = 1$)
Stein factor for general $h$

Proposition

Let $\kappa_1 = \sup_{y \in S(p)} h(y) - \inf_{y \in S(p)} h(y)$ and $\kappa_2 = \sup_{y \in S(p)} |\Delta^{-\ell} h(y)|$. If $h$ is bounded then

1. $|g(x)| \leq \kappa_1 P(x - b \ell) \bar{P}(x - b \ell) p(x + \ell)$
2. $|\mid \Delta - \ell g(x) \mid | \leq \kappa_1 c(x) (1 + |\tau \ell p c(x)| c(x + \ell)) P(x - b \ell) \bar{P}(x - b \ell) p(x + \ell)$.

If, moreover, $c = -L \ell p \eta$, three other bounds are provided.

Some references for $\ell = 0$: Bödeker (2012, 2015).
Stein factor for general $h$

**Proposition**

Let $\kappa_1 = \sup_{y \in S(p)} h(y) - \inf_{y \in S(p)} h(y)$ and $\kappa_2 = \sup_{y \in S(p)} |\Delta^{-\ell} h(y)|$.

1. If $h$ is bounded then

   - $|g(x)| \leq \kappa_1 \frac{P(x-b_\ell) \bar{P}(x-b_\ell)}{p(x+\ell)} \frac{1}{c(x+\ell)}$

   - $|\Delta^{-\ell} g(x)| \leq \kappa_1 \frac{1}{c(x)} \left( 1 + \frac{|T_\ell^p c(x)|}{c(x+\ell)} \frac{P(x-b_\ell) \bar{P}(x-b_\ell)}{p(x+\ell)} \right)$.
Stein factor for general $h$

Proposition

Let $\kappa_1 = \sup_{y \in S(p)} h(y) - \inf_{y \in S(p)} h(y)$ and $\kappa_2 = \sup_{y \in S(p)} |\Delta^{-\ell} h(y)|$.

1. If $h$ is bounded then
   - $|g(x)| \leq \kappa_1 \frac{P(x-b_\ell)\tilde{P}(x-b_\ell)}{p(x+\ell)} \frac{1}{c(x+\ell)}$
   - $|\Delta^{-\ell} g(x)| \leq \kappa_1 \frac{1}{c(x)} \left( 1 + \frac{|T_p^\ell c(x)|}{c(x+\ell)} \frac{P(x-b_\ell)\tilde{P}(x-b_\ell)}{p(x+\ell)} \right)$.

2. If $\Delta^{-\ell} h$ exists and is bounded then
   - $|g(x)| \leq \kappa_2 \frac{\tau_p^\ell(x+\ell)}{c(x+\ell)}$
   - $|\Delta^{-\ell} g(x)| \leq \kappa_2 \left( \frac{|x-\mathbb{E}[X]|}{c(x)} + \frac{|T_p^\ell c(x)|}{c(x)} \frac{\tau_p^\ell(x+\ell)}{c(x+\ell)} \right)$. 

If, moreover, $c = -L\ell p \eta$, three other bounds are provided. Some references for $\ell = 0$: Döbler (2012, 2015).
Stein factor for general $h$

Proposition

Let $\kappa_1 = \sup_{y \in S(p)} h(y) - \inf_{y \in S(p)} h(y)$ and $\kappa_2 = \sup_{y \in S(p)} |\Delta^{-\ell} h(y)|$.

1. If $h$ is bounded then
   - $|g(x)| \leq \kappa_1 \frac{P(x-b_\ell)\bar{P}(x-b_\ell)}{p(x+\ell)} \frac{1}{c(x+\ell)}$
   - $|\Delta^{-\ell} g(x)| \leq \kappa_1 \frac{1}{c(x)} \left( 1 + \frac{|T_p\ell c(x)|}{c(x+\ell)} \frac{P(x-b_\ell)\bar{P}(x-b_\ell)}{p(x+\ell)} \right)$.

2. If $\Delta^{-\ell} h$ exists and is bounded then
   - $|g(x)| \leq \kappa_2 \frac{T_p\ell(x+\ell)}{c(x+\ell)}$
   - $|\Delta^{-\ell} g(x)| \leq \kappa_2 \left( \frac{|x-E[X]|}{c(x)} + \frac{|T_p\ell c(x)|}{c(x)} \frac{T_p\ell(x+\ell)}{c(x+\ell)} \right)$.

If, moreover, $c = -L_p\eta$, three other bounds are provided.
Stein factor for general $h$

**Proposition**

Let $\kappa_1 = \sup_{y \in S(p)} h(y) - \inf_{y \in S(p)} h(y)$ and $\kappa_2 = \sup_{y \in S(p)} |\Delta^{-\ell} h(y)|$. 

1. If $h$ is bounded then
   - $|g(x)| \leq \kappa_1 \frac{P(x-b_\ell) \bar{P}(x-b_\ell)}{p(x+\ell)} \frac{1}{c(x+\ell)}$
   - $|\Delta^{-\ell} g(x)| \leq \kappa_1 \frac{1}{c(x)} \left(1 + \frac{|T_p c(x)|}{c(x+\ell)} \frac{P(x-b_\ell) \bar{P}(x-b_\ell)}{p(x+\ell)}\right)$.

2. If $\Delta^{-\ell} h$ exists and is bounded then
   - $|g(x)| \leq \kappa_2 \frac{\tau_p(x+\ell)}{c(x+\ell)}$
   - $|\Delta^{-\ell} g(x)| \leq \kappa_2 \left(\frac{|x - \mathbb{E}[X]|}{c(x)} + \frac{|T_p c(x)|}{c(x)} \frac{\tau_p(x+\ell)}{c(x+\ell)}\right)$.

If, moreover, $c = -\mathcal{L}_p \eta$, three other bounds are provided.

**Some references for $\ell = 0$:** Döbler (2012, 2015)
Application to specific distributions

Gaussian distribution

\[ |g(x)| \leq \kappa_1 \frac{\Phi(x)(1 - \Phi(x))}{\varphi(x)} \]
\[ \leq \kappa_1 \frac{1}{2} \sqrt{\frac{\pi}{2}} \]

\[ |g'(x)| \leq \kappa_1 \left( 1 + |x| \frac{\Phi(x)(1 - \Phi(x))}{\varphi(x)} \right) \]
\[ \leq 2\kappa_1 \]
Application to specific distributions

Poisson distribution with parameter 10

Bounds on $|g^-(x)|$

Bounds on $|\Delta^+ g^-(x)|$
Bounds on IPM’s

Theorem: Stein discrepancies

Let $X_n \sim p_n$ be some random variable and let $X_\infty$ have canonical
Stein operators $T_\infty^\ell$ and $L_\infty^\ell$. Then,

$$
E h(X_n) - E h(X_\infty) = E \left[ (\eta_1(X_n) - E[\eta_1(X_\infty)]) \frac{L_\infty^\ell h(X_n + \ell_\infty)}{L_\infty^\ell \eta(X_n + \ell_\infty)} \right] \\
+ E \left[ L_\infty^\ell \eta(X_n) \Delta^{-\ell_\infty} \left( \frac{L_\infty^\ell h(\cdot + \ell_\infty)}{L_\infty^\ell \eta(\cdot + \ell_\infty)} \right) (X_n) \right] \\
= E \left[ (T_\infty^\ell c_1(X_n)) \frac{L_\infty^\ell h(X_n + \ell_\infty)}{c_1(X_n + \ell_\infty)} \right] \\
+ E \left[ c_1(X_n) \Delta^{-\ell_\infty} \left( \frac{L_\infty^\ell h(\cdot + \ell_\infty)}{c_1(\cdot + \ell_\infty)} \right) (X_n) \right].
$$

In particular, IPM can be written as suprema of either of the above.
Example: Gaussian target

\[ X_\infty \sim \mathcal{N}(0, 1) \]

Total variation distance

\[
TV(X_n, X_\infty) \leq \frac{1}{2} \sqrt{\frac{\pi}{2}} \mathbb{E} \left[ |X_n + \rho_n(X_n)| \right] + \sup_z \kappa_1^\star(z)
\]
Example: Gaussian target

\( X_\infty \sim \mathcal{N}(0, 1) \)

Total variation distance

\[
TV(X_n, X_\infty) \leq \frac{1}{2} \sqrt{\frac{\pi}{2}} \mathbb{E} \left[ |X_n + \rho_n(X_n)| \right] + \sup_z \kappa_1^*(z)
\]

Normal vs Student \( (X_n \sim t_n) \)

\[
TV(X_n, X_\infty) \leq \mathbb{E} \left[ |X_n| \left| \frac{X_n^2 - 1}{X_n^2 + n} \right| \frac{\Phi(X_n) \bar{\Phi}(X_n)}{\varphi(X_n)} \right] \leq \frac{2/\sqrt{e} - 1/2}{n - 1}
\]
Example: Gaussian target
\[ X_\infty \sim \mathcal{N}(0, 1) \]

Wasserstein distance

\[
\operatorname{Wass}(X_n, X_\infty) \leq \mathbb{E} \left[ |\rho_n(X_n) + X_n| \right] + \sup_{h \in \text{Lip}(1)} |\kappa_1^*(h)|
\]
Example: Gaussian target

\[ X_\infty \sim \mathcal{N}(0,1) \]

Wasserstein distance

\[
Wass(X_n, X_\infty) \leq \mathbb{E} \left[ |\rho_n(X_n) + X_n| \right] + \sup_{h \in \text{Lip}(1)} |\kappa_1^*(h)|
\]

Normal vs Binomial

If \( X_n \) is standardized binomial with parameters \((n, \theta)\),

\[
Wass(X_n, X_\infty) \leq 2 \sqrt{\frac{1}{\theta} - 1} \frac{1}{\sqrt{n}} + \frac{2}{\sqrt{n\theta(1-\theta)}} + (1 - \theta)^n
\]
Example: Beta vs Gamma

TV between Beta($a, 3$) and Gamma($a, a + 3$)

We can choose $X_\infty \sim \text{Beta}(a, 3)$ and $X_n \sim \text{Gamma}(a, a + 3)$ or the opposite.
Part 1: spatial data

Conclusion

Part 1: spatial data

• Ernst, M. and G. Haesbroeck. Robustness of tests for spatial autocorrelation.

Part 2: Stein’s method

Thank you